

The Plan-a-Day Approach to Measuring Planning Ability in Patients with Schizophrenia

Daniel V. Holt,¹ Katlehn Rodewald,² Mirjam Rentrop,² Joachim Funke,¹ Matthias Weisbrod,^{2,3}
AND Stefan Kaiser^{2,4}

¹Department of Psychology, University of Heidelberg, Heidelberg, Germany

²Section of Experimental Psychopathology, Department of Psychiatry, University Hospital Heidelberg, Heidelberg, Germany

³Department of Psychiatry, SRH Klinikum Karlsbad-Langensteinbach, Karlsbad-Langensteinbach, Germany

⁴Psychiatric University Hospital Zurich, Zurich, Switzerland

(RECEIVED July 30, 2010; FINAL REVISION December 15, 2010; ACCEPTED December 15, 2010)

Abstract

Deficits in executive functioning are closely related to the level of everyday functioning in patients with schizophrenia. However, many existing neuropsychological measures are limited in their ability to predict functional outcome. To contribute towards closing this gap, we developed a computer-based test of planning ability (“Plan-a-Day”) that requires participants to create daily activity schedules in a simulated work setting. Eighty patients diagnosed with schizophrenia were tested with Plan-a-Day and a battery of cognitive ability tests. Plan-a-Day showed satisfactory psychometric properties in terms of consistency, reliability, and construct validity. Compared to other neuropsychological tests used in this study, it also demonstrated incremental validity with regard to the Global Assessment of Functioning. The Plan-a-Day approach, therefore, seems to represent a valid alternative for measuring planning ability in patients with executive function deficits, occupying a middle ground between traditional neuropsychological tests and real-life assessments. (*JINS*, 2011, *17*, 327–335)

Keywords: Executive functioning, Planning, Cognitive abilities, Computer-based testing, Schizophrenia, Neuropsychological testing, Ecological validity

INTRODUCTION

Past research has shown that executive functions play an important role in predicting everyday functioning in psychiatric patients (Green, 1996; Green, Kern, Braff, & Mintz, 2000; Jaeger & Douglas, 1992; Velligan, Bow-Thomas, Mahurin, Miller, & Halgunseth, 2000). From a diagnostic and therapeutic perspective, the ability to measure executive functioning reliably and in an ecologically valid manner, therefore, is a relevant issue. This article focuses on how the measurement of planning ability can be improved by using the computer-based test “Plan-a-Day,” which simulates day planning scenarios in a work setting. We present data from patients with schizophrenia, but propose that the test can also be used in other populations.

Deficits in executive functioning in general and planning difficulties in particular are commonly observed in patients with schizophrenia (e.g., Morris, Rushe, Woodruffe, & Murray, 1995; Pantelis et al., 1997). A possible explanation

for this could be difficulties in translating plans into “willed intentions,” which may be related to impaired frontostriatal connectivity (Frith, 1987; Robbins, 1990). More recently, Burgess, Dumontheil, and Gilbert (2007) suggested a neuro-cognitive explanation that may underlie several executive deficits and seems particularly applicable to planning difficulties. According to their gateway hypothesis of rostral prefrontal cortex function, the ability to switch between external stimulus-oriented representations (e.g., the task at hand) and internal stimulus-independent thought (e.g., creative thoughts, plans) may be impaired in patients with executive deficits. As planning ability is highly relevant for many everyday activities such as shopping, cooking, or successfully participating in the job market, valid measurement of this ability seems desirable from a diagnostic point of view and may be useful for psychiatric rehabilitation.

ECOLOGICAL VALIDITY OF TESTS OF EXECUTIVE FUNCTIONING

The question of how to improve the ecological validity of neuropsychological tests of executive functioning is a matter

Correspondence and reprint requests to: Daniel V. Holt, Department of Psychology, University of Heidelberg, Hauptstr. 47-51, 69115 Heidelberg, Germany. E-mail: daniel.holt@psychologie.uni-heidelberg.de

of ongoing debate (e.g., Burgess et al., 2006; Chaytor & Schmitter-Edgecombe, 2003; Wilson, Evans, Emslie, Alderman, & Burgess, 1998). In this context, ecological validity can be defined as “the degree to which test performance corresponds to real-world performance” (Chaytor & Schmitter-Edgecombe, 2003, p. 182). Many existing tests of executive functioning were originally used as experimental laboratory tasks and only later repurposed as neuropsychological diagnostics, for example the Stroop task or the Tower of London and its variants (Burgess et al., 2006). These tests are easy to administer and score, have reasonable psychometric reliability and show satisfactory convergent validity with respect to other standard tests of cognitive ability. However, their relation to real-life outcome measures, such as work performance or activities of daily living is often unclear or lacking (Burgess, Alderman, Volle, Benoit, & Gilbert, 2009; Manchester, Priestley, & Jackson, 2004; Shallice & Burgess, 1991). This is not entirely surprising, considering how little many standard neuropsychological tests resemble the challenges and complexities of everyday life. When the main purpose of neuropsychological testing is to assist with the diagnosis of brain pathology, this is not necessarily a problem. However, with a rising interest in ecological validity and predicting functional outcome (Chaytor & Schmitter-Edgecombe, 2003; Manchester et al., 2004), it becomes pertinent to ask to what extent existing neuropsychological test procedures possess ecological validity and how it can be increased (Chaytor, Schmitter-Edgecombe, & Burr, 2006). This has also been termed the “veridicality approach” to ecological validity (Chaytor & Schmitter-Edgecombe, 2003; Chaytor et al., 2006). It can be contrasted with the “verisimilitude approach” that involves constructing new testing paradigms to capture the cognitive demands of real-life tasks more accurately. Burgess and colleagues (2006) have suggested several task demands that pose difficulties for patients with impaired executive functioning, yet are absent in many traditional neuropsychological tests. Among these are the ability to multi-task, work for comparatively long periods of time on a task without receiving feedback, and, more generally, to find a goal-oriented and structured approach to “ill structured” and complex tasks (Burgess, 2000; Goel, Grafman, Tajik, Gana, & Danto, 1997).

The Behavioral Assessment of the Dysexecutive Syndrome (BADS; Wilson, Alderman, Burgess, Emslie, & Evans, 1996), which compiles several tests of executive functioning into one standardized test battery, has been developed to meet these requirements using an office-based test format. Alternatively, there are real-life assessments of activities of daily living, such as shopping, cooking, or running errands (e.g., Knight, Alderman, & Burgess, 2002; Rempfer, Hamera, Brown, & Cromwell, 2003; Semkovska, Bédard, Godbout, Limoge, & Stip, 2004; Shallice & Burgess, 1991), which emphasize maximum realism at the expense of tying the test to local conditions, such as the availability of suitable shopping opportunities. In between classical office-based tests and real-life assessments lie computer-based simulation tests that provide a certain degree of realism and complexity while maintaining high control over the task. Some computer-based tests use

realistic three-dimensional virtual reality environments (e.g., Kurtz, Baker, Pearlson, & Astur, 2007; McGeorge et al., 2001; Rand, Rukan, Weiss, & Katz, 2009), whereas others consist of schematic representations of real-life situations, simplified with respect to visual presentation and possibilities for interaction (e.g., Craik & Bialystok, 2006; Larøi, Canlaire, Mourad, & Van Der Linden, 2010). While the former approach has an appeal in terms of approximating real life as closely as possible, the latter encourages a focus on task characteristics essential for assessment purposes as opposed to realistic but potentially incidental surface features. The Plan-a-Day test belongs to this category.

It is important to note that, while high realism usually affords face validity, it does not guarantee ecological validity. More important than surface resemblance to real-life situations is whether a test taps the ability (or combination of abilities) relevant for particular outcome situations. Therefore, the ecological validity of neuropsychological tests should be shown empirically, regardless of their apparent degree of realism (cf. Chaytor & Schmitter-Edgecombe, 2003). Furthermore, tests with a high degree of realism may also be less specific with respect to the cognitive constructs they measure, as most real-life activities require a combination of several cognitive abilities.

THE PLAN-A-DAY APPROACH TO MEASURING PLANNING ABILITY

Plan-a-Day is a computer-based scheduling task framed in workplace semantics that requires participants to schedule a list of work-related activities (e.g., picking up mail at the post office or checking inventory at a warehouse) while considering various constraints about when, where, and for what duration the activities have to be carried out. The difficulty of each Plan-a-Day problem is determined by the number of tasks to be scheduled and the number and interaction of constraints that need to be considered. In its present form Plan-a-Day is a modified version of an initial design developed for a job assessment center, which had already been tested with healthy control samples (Funke & Krüger, 1995). Results indicated moderate levels of reliability and provided preliminary evidence for criterion validity in a context of personnel selection. The present version of the test was modified to increase measurement reliability and adapt task difficulty for patients with mild to moderate impairments of executive functioning. We decided to carry out the validation study with a sample of patients with schizophrenia that had a comparatively high level of functioning. This group optimally matched the test profile, both with respect to general executive functioning and the level of planning ability in particular. However, we assume that findings from this group are likely to generalize to other patient populations with a comparable level of executive impairment.

Many existing neuropsychological planning tests, such as the Tower of London or labyrinth tasks, are move planning tasks with a strong visuo-spatial component. Plan-a-Day in contrast is a scheduling task that requires planning in the

temporal domain, drawing mostly on verbally encoded information. As such, it complements existing tests by covering another planning domain and a different mode of representation. Moreover, temporal scheduling problems are a common feature of everyday life, whereas most move planning problems do not have direct real-life equivalents.

The formal tractability of the type of constraint-satisfaction problem used in Plan-a-Day allows the systematic construction of items with varying degrees of complexity and hence difficulty. At the same time, the parameter space available for item construction is comparatively large, so that each item can pose its own slightly different challenges, requiring the flexible use and monitoring of various heuristics and strategies for finding the solution. Workplace semantics were chosen to increase the face validity of the task, which may be important for acceptance of the test and, therefore, for motivation and engagement of participants. Furthermore, the semantics also help to activate pre-existing knowledge about approaching scheduling problems (cf. Blessing & Ross, 1996), which is desirable from a perspective of ecologically valid testing. Finally, the computer implementation of Plan-a-Day allows an easy, economical, and standardized administration of the test with automatic data logging and scoring.

Aims and Hypotheses

The purpose of the present study was to validate Plan-a-Day in a sample of patients with schizophrenia entering a rehabilitation program in preparation for returning to a work environment. Our hypotheses were as follows: (1) Plan-a-Day shows good internal consistency and acceptable retest-reliability. (2) Plan-a-Day performance correlates with performance in other planning tests indicating good construct validity and specificity. (3) Compared to standard neuropsychological tests, Plan-a-Day can explain additional variance on a global level of functioning, thus demonstrating ecological and incremental validity.

METHOD

Design

This study was carried out as part of a project comparing the effects of planning and problem solving training with the training of basic cognitive functions (processing speed, attention, memory) in cognitive rehabilitation. A neuropsychological test battery including the Plan-a-Day test was administered twice with a 4-week interval during which participants received cognitive ability training and inpatient work therapy. For introducing the Plan-a-Day test, we will focus on data from the first measurement.

Participants

Eighty patients meeting Diagnostic and Statistical Manual of Mental Disorders—Fourth Edition (DSM-IV) criteria for schizophrenia or schizoaffective disorder participated in the study. The diagnosis was confirmed through the Mini-International

Table 1. Demographic and clinical characteristics of patients included in the study ($N = 80$)

Categorical variables	<i>N</i>	%
Gender		
Male	63	78.8
Female	17	21.2
Diagnoses (DSM-IV)		
Schizophrenia, paranoid	60	75
Schizophrenia, disorganized	1	1.3
Schizophrenia, residual	2	2.5
Schizoaffective disorder	15	18.8
Schizophrenia simplex	2	2.5
Continuous variables	Mean	<i>SD</i>
Age	29.60	8.31
Years of formal education	15.28	3.77
Estimated premorbid IQ (MWT-B) ^a	104	13.64
Age at first hospitalization	25.08	7.76
GAF	59.90	6.53
PANSS		
Positive	12.49	2.88
Negative	18.70	4.14
Global	31.91	6.22
Total	63.10	10.64

Note. DSM-IV = Diagnostic and Statistical Manual of Mental Disorders—Fourth Edition; MWT-B = Mehrfachwahl-Wortschatzintelligenz-Test Version B; GAF = Global Assessment of Functioning; PANSS = Positive and Negative Syndrome Scale.

^aMWT-B raw scores: $M = 27.17$ ($SD = 4.89$); for this test $N = 77$.

Neuropsychiatric Interview (MINI; Ackenheil, Dietz-Bauer, & Vossen, 1998). Participants were recruited from an inpatient unit at the mental health hospital SRH Klinik Karlsbad-Langensteinbach, Germany. Patients were living in the community before entering a treatment program aimed at facilitating return to work. This included patients with persistent problems after an acute illness episode as well as those with a longer illness course. In addition to a diagnosis of schizophrenia or schizoaffective disorder inclusion criteria were (1) age of 18 or older, (2) being in a post-acute phase of illness (score of all Positive and Negative Syndrome Scale [PANSS] positive items < 5), and (3) having an estimated premorbid IQ of 80 or above. Exclusion criteria included (1) a primary diagnosis of a neurological disorder, (2) illicit substance use during the last month, and (3) having a current comorbid Axis I disorder. All participants gave written informed consent, and the study was approved by the ethics committee of the University of Heidelberg Medical Faculty. For demographic and clinical characteristics of participants see Table 1.

Plan-a-Day

Plan-a-Day¹ is a computer-based scheduling task with workplace semantics. Participants are asked to imagine that

¹ English and German versions of the test, including instructions and a technical manual, are available from the website of the corresponding author (<http://www.atp.uni-hd.de/tools/planaday>).

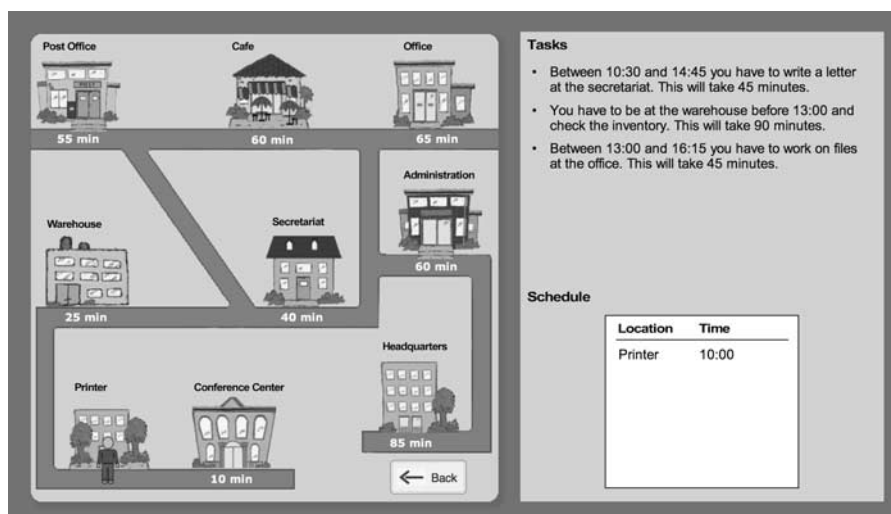


Fig. 1. Plan-a-Day screenshot at the start of a problem. The current location, indicated by a small manikin on the map, is at the “Printer.” Below each location the travel time from the current location is displayed. Tasks to be scheduled are displayed in the top right hand corner, the planned schedule for this day (bottom right) so far only contains the starting location and current time.

they work for a small company where they have to plan their daily activities. Information about the tasks to be carried out each day is presented in a task information area on the right side of the screen (see Figure 1), while the left side of the screen displays the different locations, distances between locations, and current position. The constraints that need to be considered for solving Plan-a-Day problems include earliest start, latest finish, location, duration of tasks, and the distance between different locations. Participants are instructed to first plan their daily activities and then implement their plan by moving the symbol representing their current positions by clicking on the corresponding locations on the map. If they are too late to carry out a task at a location, a small information window with a corresponding message is displayed. There is only one correct solution for each Plan-a-Day problem. When participants notice a mistake in their plan, they can undo previous moves using the “back”-button, which is accompanied by an information window confirming the undo operation and an acoustic signal. Participants are instructed to avoid using the “back”-button by planning ahead appropriately.

Assessment problems were designed by systematically varying two dimensions of working memory load and computational complexity: the number of errands, ranging from two to four per problem, and the number of information elements that need to be considered to solve the problem. The second dimension had three levels: problems that can be solved just by looking at start and end times (i.e., there is no overlap of the time frames for different errands), problems that also require considering errand durations, and problems that require errand durations and the time needed for reaching a particular location. Figure 2 displays the structure of a typical Plan-a-Day problem with four tasks and a medium difficulty level.

The test consisted of two practice problems and eight assessment problems of increasing difficulty, requiring 20 to 30 minutes for completion. The operation of the program was

explained and interactively demonstrated with the first practice problem, using scripted instructions. Participants were then given the opportunity to further practice the operation of the program with the second practice problem. No time limit was set for the main assessment phase but the test program could be exited early after six of eight assessment problems if a participant seemed overchallenged.

Three scores were calculated for the Plan-a-Day test: total solution time, the planning ratio (percentage of total time spent planning), and the number of problems solved without corrections (i.e., not using the “back”-button during the execution phase). Plan-a-Day was designed in such a manner that solution time captures most of the performance-relevant information. To achieve this, the program only proceeded to a new problem when the previous problem had been correctly solved. As a side effect, this may have improved participant motivation as every task is ultimately successfully solved. This mechanism also means that planning mistakes incur a time penalty: Undoing an incorrect move requires clicking the “back”-button and confirming the corresponding information message with another click before the move is undone. The number of solutions involving no corrections was analyzed as it may be comparable to the accuracy score used in other tests, and the planning ratio seemed a promising measure for how well-planned and strategic participants proceeded.

Neuropsychological Tests

In addition to Plan-a-Day, planning ability was measured with a computer-based variant of the Tower of London (Kohler & Beck, 2004) using eight planning problems of increasing difficulty, and the Zoo-Map subtest from the German version of the BADS (Ufer, 2000). The Zoo-Map test from the BADS requires participants to plan how they would visit a range of locations at a zoo considering a given

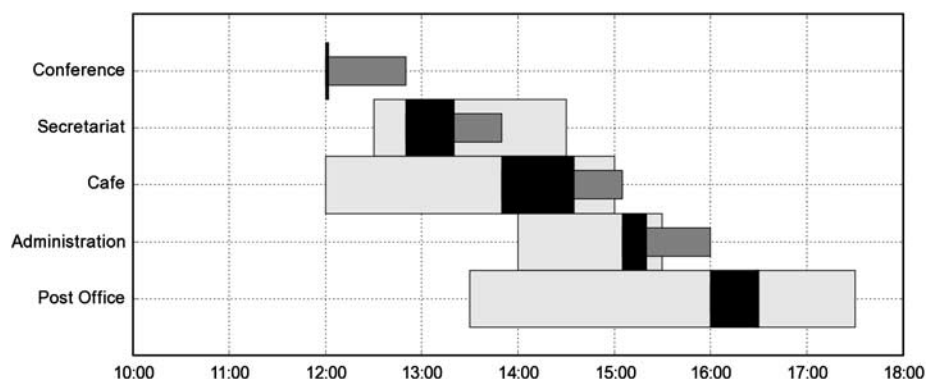


Fig. 2. Schematic display of the structure of a Plan-a-Day item with four tasks and medium computational complexity. Locations to be visited are displayed on the vertical axis, descending in order of the correct solution. Virtual time is displayed on the horizontal axis. Large light gray bars indicate the time window during which a task may be carried out at each location. The black line at the top marks the starting point (“Conference”), thin dark gray bars indicate travel time, black bars execution of a task at a specific location. The tasks to be scheduled for this item were:

*“Between 14:00 and 15:30 you have to go to the administration building to sign a contract. This will take 15 minutes.
Between 12:00 and 15:00 you want to meet a business partner at the cafe. The meeting will last for 45 minutes.
Between 12:30 and 14:30 you have to dictate a letter to the secretariat. This is going to take 30 minutes.
You have to pick up your mail between 13:30 and 17:30 at the post office. This will take 30 minutes.”*

set of rules and draw their solution on a map using a pen. Working memory was assessed with three subtests from the German version of the Wechsler Adult Intelligence Scale—Third Edition (von Aster, Neubauer, & Horn, 2006) to assess verbal memory maintenance and manipulation: Digit Span Forward, Digit Span Backward and Letter-Number Sequencing. The Corsi Block-Tapping Task was used to assess spatial working memory maintenance and manipulation analogous to Digit Span Forward and Backward (Schelling, 1993). The Trail Making Test (TMT; Reitan, 1992) and a simple color Stroop test (Markela-Lerenc, Kaiser, Fiedler, Weisbrod, & Mundt, 2006) were used to assess processing speed (TMT, Version A and reaction time in the Stroop neutral condition) and response inhibition (TMT Version B and difference between congruent and incongruent Stroop trials). Furthermore, a vocabulary-based test (Mehrfachwahl-Wortschatz-Intelligenztest Version B [MWT-B]; Lehrl, Triebig, & Fischer, 1995) was used to obtain an estimate of premorbid crystallized intelligence and a test of arithmetic ability (Zahlenverarbeitungs- und Rechentest [ZRT]; Kalbe, Brand, & Kessler, 2002) was also included.

Clinical and Functional Assessment

Psychiatric symptoms were assessed by interview using the Positive and Negative Syndrome Scale (Kay, Fiszbein, & Opfer, 1987), a 30-item symptom rating scale completed by clinically trained research staff.

The first measure of everyday functioning was the Global Assessment of Functioning (GAF) according to DSM-IV criteria, a single item rating on a scale from 0 to 100. The rating was given by a trained research psychologist as part of the intake-interview. The GAF rating is a commonly used measure for axis V of the DSM-IV to assess psychological,

social and work-related functioning based on descriptions of ten broad levels of functioning. Despite being only a single-item scale, the GAF shows good inter-rater reliability and satisfactory concurrent validity with other instruments measuring levels of functioning and psychosocial impairments (Hilsenroth et al., 2000; Startup, Jackson, & Bendix, 2002). Functional capacity was also assessed with the 30-item Osnabrück Work Capabilities Profile (Wiedl & Uhlhorn, 2006), which is based on the widely used Work Personality Profile (Bolton & Roessler, 1986). Using the O-AFP, a work therapist rated functional capacity based on the patients’ performance in work therapy.

RESULTS

Consistency and Reliability

Average scores for the Plan-a-Day test are displayed in Table 2. The internal consistency (Cronbach’s α) of Plan-a-Day was relatively high for total solution time with $\alpha = .78$. The planning ratio also turned out to be a moderately consistent characteristic with $\alpha = .67$, whereas the consistency of the accuracy score was rather low with $\alpha = .47$.

A retest was carried out 4 weeks after the initial assessment. In the interval, all patients participated in the regular treatment

Table 2. Descriptive statistics for Plan-a-Day performance ($N = 80$), average over all eight problems

Measure	Mean	SD	Min	Max
Solution time (s)	68.80	32.28	29	179
Planning ratio (%)	48.81	10.66	30.40	77.50
Accuracy (%)	67.97	18.43	25	100

program consisting of occupational therapy, physiotherapy, social skills training, and—for half of the sample—computer-based training of basic cognitive abilities (memory, reaction speed, continuous attention) using the RehaCom training package (Hasomed GmbH, Germany). As part of another study, the other half received a dedicated planning training instead. Test–retest reliability was only analyzed for the first group ($n = 40$) to avoid effects of the planning-specific training intervention. Still, the reported test–retest reliabilities are likely to represent lower boundaries of the actual values due to varying treatment-related gains in cognitive ability. The test–retest reliability of Plan-a-Day with an interval of 4 weeks was $r = .82$ for the solution time measure. For comparison, the reliability of the solution time for the other planning tests in this study was $r = .58$ for the Tower task and $r = .59$ for the Zoo-Map. The test–retest reliability of the planning ratio, $r = .39$, and accuracy, $r = .33$, showed that these measures were not particularly stable over time.

Construct Validity

Table 3 illustrates the relation of Plan-a-Day performance measures to other cognitive ability tests.² With a sample size of 80 participants, the study provided sufficient statistical test power to reliably detect small to medium correlations between tests with 80% test power to detect $r > .30$ at an alpha-level of .05. As expected, Plan-a-Day solution time showed clear convergent validity with other planning tests, and a small but statistically significant correlation with Stroop neutral reaction time. Discriminant validity was shown with respect to both verbal and spatial working memory, as well as crystallized intelligence as estimated by the MWT-B. The planning ratio measure, however, behaved differently and was associated with spatial working memory, arithmetic ability, and crystallized intelligence. As for accuracy, there was a small correlation with Letter-Number Sequencing and also with crystallized intelligence. Accuracy also showed a moderate correlation with years of formal education, $r = .36$, $p < .001$. None of the Plan-a-Day measures was significantly correlated with psychiatric symptoms as measured by the PANSS positive, negative, and global scales.

Ecological Validity

The raw correlations between Plan-a-Day and the GAF were $r = -.32$, $p < .01$, and $r = -.15$, not significant, between Plan-a-Day and the OAF-P. This implies that 10.2% of variance in the GAF can be explained by Plan-a-Day alone. To establish the incremental validity of Plan-a-Day, we entered the main ability variables into a stepwise regression to find the best predictors of everyday functioning for this test battery. Of all tests other than Plan-a-Day, only Letter-Number Sequencing was able to predict functional capacity as measured by the GAF to a notable extent, $R^2 = .21$,

² If non-parametric Spearman's rho correlations are used, all significance levels remain identical except for Stroop time, $\rho = .18$, $p = .18$.

Table 3. Convergent and discriminant validity of Plan-a-Day expressed as correlation (Pearson's r) with other neuropsychological tests ($N = 80$)

	Plan-a-Day		
	Solution time	Planning ratio	Accuracy
Planning			
Tower-Task Time	.42**	-.01	-.01
Tower-Task Accuracy	.10	-.24*	.12
Zoo-Map Time	.37**	.01	-.11
Zoo-Map Accuracy	-.12	.16	.19
Working Memory			
Letter-Number Sequencing	-.20	.34**	.34**
Digit Span Forward	-.04	.10	.18
Digit Span Backward	.08	.22	.14
Corsi Forward	-.06	.00	-.05
Corsi Backward	-.06	.00	-.09
Other			
TMT-A Time	.15	-.17	-.16
TMT-B Time	.20	-.11	-.15
Stroop Interference ($N = 75$)	-.19	.06	.21
Stroop Neutral Time ($N = 75$)	.23 *	.11	.08
MWT-B ($N = 77$)	-.08	.26 *	-.30**
ZRT (arithmetic ability)	.02	.23 *	.18

Note. TMT = Trail Making Test; MWT-B = Mehrfachwahl-Wortschatz-Intelligenztest Version B; ZRT = Zahlenverarbeitungs- und Rechentest;

*Statistically significant at $p < .05$.

**Statistically significant at $p < .01$.

$\beta = .42$, $p < .001$. Adding Plan-a-Day as a predictor showed that it was able to explain a significant amount of unique variance in the criterion beyond Letter-Number Sequencing, $R^2 = .26$, $\Delta R^2 = .05$, $\beta = -.24$, $p < .05$, showing its incremental validity. No other cognitive test in this study was able to significantly enhance the prediction of the GAF any further.

DISCUSSION

The present data support Plan-a-Day as a useful instrument for measuring planning ability at the level of executive functioning. The principal measure of Plan-a-Day (solution time) showed a satisfactory level of reliability, good convergence with other measures of planning ability and contributed to predicting global functioning. We will now discuss the properties of different measures derived from the Plan-a-Day test, as well as potential limitations and future extensions.

Consistency and Reliability

The internal consistency of the principal measure of Plan-a-Day (solution time) was satisfactory with an acceptable level of test–retest reliability, comparing favorably with the other planning tests used in this study. The planning ratio measure also turned out to be surprisingly consistent, although not particularly stable, while the accuracy score showed

comparatively low values in both respects. Many participants apparently followed the instruction to take a longer time planning and keep the number of errors low. The planning ratio may reflect such strategic and intentional behavior, which is conceivably more variable across different testing sessions than measures of basic cognitive capacity. The stability of this variable was probably further affected by the cognitive training interventions between test and retest. The low internal consistency of the accuracy measure may be due to a related reason: When participants deliberately spend a longer time planning to avoid errors, ability-related systematic variance is shifted away from the accuracy score toward the time-based measures, as intended by design. Since the accuracy score still seemed to capture a facet of performance notably different from solution time, it may be worthwhile to reconsider this initial design goal and investigate possibilities to make this measure more reliable.

Construct Validity

Results seem to generally support the role of Plan-a-Day as a specific planning test. Solution time showed clear associations with other measures of planning, while discriminant validity was shown with respect to most other cognitive ability tests, except for a small correlation with reaction time in the Stroop task. This may indicate a processing speed component, which seems plausible for a time-based performance measure. As mental planning is a process that clearly draws on working memory, the weak link to working memory scores seems puzzling. However, all task information is permanently displayed on screen, which may render raw working memory capacity secondary to the ability to create and monitor a working strategy in the face of a comparatively complex and novel problem. This contrasts with the demands posed by simple and clearly defined tasks such as the working memory tests used in this study. The working memory test that comes closest to reaching a meaningful correlation with Plan-a-Day solution time is the one with the most complex task set and highest demand on mental manipulation (Letter-Number Sequencing), which incidentally was the only other task that predicted global functioning to a significant extent.

Considering other Plan-a-Day variables, the planning ratio emerged not only as a reasonably consistent overall measure, but also showed a characteristic pattern of relations to Tower task accuracy, spatial working memory, arithmetic ability, and—like the accuracy score—Letter-Number Sequencing and crystallized intelligence. Assuming that the planning ratio reflects strategic behavior, it seems plausible that it correlates with tasks benefiting from a well-planned approach and *ad hoc* strategies. This could also explain why comparatively simple, less strategy-prone tasks (e.g., TMT-A, Stroop, Digit Span Forward) did not show a strong relation to this indicator. Similar reasons may explain the results for the Plan-a-Day accuracy score, which also correlated with Letter-Number Sequencing and crystallized intelligence, as a high accuracy score can be achieved through strategic and well-planned behavior.

Several critical points can be raised with respect to construct validity. First, considering the amount of arithmetic involved in solving a Plan-a-Day problem one may be inclined to think that Plan-a-Day is just a semantically framed test of mental arithmetic. Although performing simple calculations is certainly part of the task, the near zero correlation of Plan-a-Day and the arithmetic test shows that this is unlikely to be the dominating element. Plan-a-Day may be less characterized by mathematical complexity (simple addition will do), but rather by the necessity to extract and integrate relevant information from the task environment.

Second, we have claimed that Plan-a-Day does not have a strong visuo-spatial component, but does the prominent map displayed on screen not directly contradict this statement? The map may be helpful for participants to more vividly visualize the simulated task setting; however, there are reasons to believe this does not constitute a strong visuo-spatial component. All information is presented verbally and numerically on screen and the spatial layout of the map alone is neither required nor sufficiently precise to solve the problems. This was supported empirically by the discriminant validity of Plan-a-Day performance and the Corsi Block Tapping test of spatial working memory.

One important omission in the present test battery is a test of fluid intelligence, which was not included for time reasons. It is possible that fluid intelligence may explain a significant amount of variance in the Plan-a-Day task, as it is correlated with problem solving ability in many domains. This gap may be filled by future studies.

Ecological Validity

Besides largely showing the expected pattern in terms of construct validity, Plan-a-Day also demonstrated incremental validity in predicting global functioning as measured by the GAF. This is an encouraging result, as one aim of the development of Plan-a-Day was to move closer to measuring real-life functioning, for which the GAF is a first approximation. These results should be viewed with some caution, as the O-AFP measure of functional capacity was not significantly correlated with Plan-a-Day variables. However, the O-AFP did not show systematic relations to any of the ability tests used in this study, although its basic psychometric properties have been shown to be adequate in a large norming study (Wiedl & Uhlhorn, 2006). It seems plausible that the OAF-P captures an aspect of functional capacity not directly related to the cognitive functions measured by the test battery used in the present study, emphasizing higher-level social and organizational skills instead (e.g., learning from feedback, punctuality, adequate communication).

There may be several reasons why Plan-a-Day was able to deliver an increment in external validity above other neuropsychological tests. The realistic task setting may play a role in increasing the transferability of test results to everyday situations, or conversely, activate existing prior knowledge and skills relevant for approaching the task. Additionally, the apparent face validity may also motivate participants to take

the task seriously as they see its relation to similar real-life situations. From a different perspective, following the reasoning of Burgess et al. (2006), what may lie at the heart of Plan-a-Day as an executive function test is a realistic degree of complexity beyond mere surface semantics as well as its comparative novelty as a task. It requires a considerable amount of strategic thinking, flexible adaptation of prior knowledge, and meta-cognition to develop and monitor a working strategy for solving Plan-a-Day problems. In this respect, it fulfills several of the requirements that have been put forth as desirable for ecologically valid measures of executive functioning (e.g., Burgess, 2000; Goel et al., 1997): Participants have to develop a strategy to handle a comparatively complex and “ill structured” situation and do not receive direct feedback while working on the task for a relatively long period of time. While Plan-a-Day does not require multi-tasking in a narrow sense (participants only work on the scheduling task), it does require setting and balancing priorities of the tasks to be scheduled.

Plan-a-Day as an Instrument for Clinical Assessment and Research

The present results show that Plan-a-Day is a valid and reliable instrument for measuring planning ability, with some indication that it also possesses incremental validity with regard to everyday functioning. Although further studies are needed to verify and extend the present findings, these results render Plan-a-Day an interesting option for neuropsychological assessment. Additionally, Plan-a-Day is easy to administer and score, and we believe that the face validity afforded by the workplace semantics has a positive influence on motivation and acceptance of the test on part of the participants.

Further studies that use a wider range of functional outcome measures and investigate the ability of Plan-a-Day to distinguish between relevant criterion groups (e.g., patients with executive deficits versus healthy controls) are desirable. Another priority will be to exactly quantify the task parameters that determine difficulty and differential external validity of Plan-a-Day problems. Furthermore, a short version of the test and test norms for use in individual assessment are currently under development.

Plan-a-Day may also offer some interesting perspectives for basic research. In contrast to the largely visuo-spatial movement planning problems often used in planning research, it constitutes a largely verbally presented temporal planning problem. This could be an advantage for studying modality-independent aspects of planning by contrasting Plan-a-Day with other paradigms. At the same time the task can easily be formally described, which enables systematic analysis and manipulation of various task parameters for experimental studies.

In summary, it appears that the Plan-a-Day approach shows some promise to fill the middle ground between traditional neuropsychological tests and real-world tasks, while offering interesting perspectives for both clinical application and research.

ACKNOWLEDGMENTS

This research was supported by the German Federal Ministry of Education and Research (Bundesministerium für Bildung und Forschung), grant number 01GW0639. D.H. and J.F. have developed a commercially distributed computer-based training for planning ability, which is based on principles similar to the test described in this article. We thank two anonymous reviewers for their helpful comments.

REFERENCES

- Ackenheil, M., Dietz-Bauer, R., & Vossen, A. (1998). *Mini international neuropsychiatric interview, German version 5.0.0, DSM IV*. München, Germany: Psychiatrische Universitätsklinik München.
- Blessing, S.B., & Ross, B.H. (1996). Content effects in problem categorization and problem solving. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 22, 792–810.
- Bolton, B., & Roessler, R. (1986). The work personality profile: Factor scales, reliability, validity, and norms. *Vocational Evaluation & Work Adjustment Bulletin*, 19, 143–149.
- Burgess, P.W. (2000). Strategy application disorder: The role of the frontal lobes in human multitasking. *Psychological Research*, 63, 279–288.
- Burgess, P.W., Alderman, N., Forbes, C., Costello, A., Coates, L.M.-A., Dawson, D.R., ... Channon, S. (2006). The case for the development and use of ‘ecologically valid’ measures of executive function in experimental and clinical neuropsychology. *Journal of the International Neuropsychological Society*, 12, 194–209.
- Burgess, P.W., Alderman, N., Volle, E., Benoit, R.G., & Gilbert, S.J. (2009). Mesulam’s frontal lobe mystery re-examined. *Restorative Neurology and Neuroscience*, 27, 493–506.
- Burgess, P.W., Dumontheil, I., & Gilbert, S.J. (2007). The gateway hypothesis of rostral prefrontal cortex (area 10) function. *Trends in Cognitive Sciences*, 11, 290–298.
- Chaytor, N., & Schmitter-Edgecombe, M. (2003). The ecological validity of neuropsychological tests: A review of the literature on everyday cognitive skills. *Neuropsychology Review*, 13, 181–197.
- Chaytor, N., Schmitter-Edgecombe, M., & Burr, R. (2006). Improving the ecological validity of executive functioning assessment. *Archives of Clinical Neuropsychology*, 21, 217–227.
- Craik, F.I.M., & Bialystok, E. (2006). Planning and task management in older adults: Cooking breakfast. *Memory & Cognition*, 34, 1236–1249.
- Frith, C.D. (1987). The positive and negative symptoms of schizophrenia reflect impairments in the perception and initiation of action. *Psychological Medicine*, 17, 631–648.
- Funke, J., & Krüger, T. (1995). “Plan-A-Day”: Konzeption eines modifizierbaren Instruments zur Führungskräfte-Auswahl sowie erste empirische Befunde [“Plan-A-Day”: Design of a modifiable instrument for selecting managerial personnel and first empirical results]. In J. Funke, T. Krüger, & A. Fritz (Eds.), *Neue Konzepte und Instrumente zur Planungsdiagnostik* (pp. 97–120). Bonn, Germany: Deutscher Psychologen Verlag.
- Goel, V., Grafman, J., Tajik, J., Gana, S., & Danto, D. (1997). A study of the performance of patients with frontal lobe lesions in a financial planning task. *Brain: A Journal of Neurology*, 120, 1805–1822.

- Green, M.F. (1996). What are the functional consequences of neurocognitive deficits in schizophrenia? *The American Journal of Psychiatry*, 153, 321–330.
- Green, M.F., Kern, R.S., Braff, D.L., & Mintz, J. (2000). Neurocognitive deficits and functional outcome in schizophrenia: Are we measuring the “right stuff”? *Schizophrenia Bulletin*, 26, 119–136.
- Hilsenroth, M.J., Ackerman, S.J., Blagys, M.D., Baumann, B.D., Baity, M.R., Smith, S.R., ... Daniel, J.J. (2000). Reliability and validity of DSM-IV axis V. *The American Journal of Psychiatry*, 157, 1858–1863.
- Jaeger, J., & Douglas, E. (1992). Neuropsychiatric rehabilitation for persistent mental illness. *Psychiatric Quarterly*, 63, 71–94.
- Kalbe, E., Brand, M., & Kessler, J. (2002). *Zahlenverarbeitungs- und Rechentest (ZRT)* [Number processing and arithmetic test (ZRT)] Weinheim, Germany: Beltz Verlag.
- Kay, S.R., Fiszbein, A., & Opfer, L.A. (1987). The Positive and Negative Syndrome Scale (PANSS) for schizophrenia. *Schizophrenia Bulletin*, 13, 261–276.
- Knight, C., Alderman, N., & Burgess, P.W. (2002). Development of a simplified version of the multiple errands test for use in hospital settings. *Neuropsychological Rehabilitation*, 12, 231–255.
- Kohler, J., & Beck, U. (2004). *Planungstest*. Konstanz, Germany: Kohler & Beck GbR.
- Kurtz, M.M., Baker, E., Pearlson, G.D., & Astur, R.S. (2007). A virtual reality apartment as a measure of medication management skills in patients with schizophrenia: A pilot study. *Schizophrenia Bulletin*, 33, 1162–1170.
- Larøi, F., Canlaire, J., Mourad, H., & Van Der Linden, M. (2010). Relations between a computerized shopping task and cognitive tests in a group of persons diagnosed with schizophrenia compared with healthy controls. *Journal of the International Neuropsychological Society*, 16, 180–189.
- Lehrl, S., Triebig, G., & Fischer, B. (1995). Multiple choice vocabulary test MWT as a valid and short test to estimate premorbid intelligence. *Acta Neurologica Scandinavica*, 91, 335–345.
- Manchester, D., Priestley, N., & Jackson, H. (2004). The assessment of executive functions: Coming out of the office. *Brain Injury*, 18, 1067–1081.
- Markela-Lerenc, J., Kaiser, S., Fiedler, P., Weisbrod, M., & Mundt, C. (2006). Stroop performance in depressive patients: A preliminary report. *Journal of Affective Disorders*, 94, 261–267.
- McGeorge, P., Phillips, L.H., Crawford, J.R., Garden, S.E., Della Sala, S., Milne, A.B., ... Callender, J.S. (2001). Using virtual environments in the assessment of executive dysfunction. *Presence: Teleoperators and Virtual Environments*, 10, 375–383.
- Morris, R.G., Rushe, T., Woodruffe, P.W.R., & Murray, R.M. (1995). Problem solving in schizophrenia: A specific deficit in planning ability. *Schizophrenia Research*, 14, 235–246.
- Pantelis, C., Barnes, T.R., Nelson, H.E., Tanner, S., Weatherley, L., Owen, A.M., & Robbins, T.W. (1997). Frontal-striatal cognitive deficits in patients with chronic schizophrenia. *Brain*, 120, 1823–1843.
- Rand, D., Rukan, S.B.A., Weiss, P.L., & Katz, N. (2009). Validation of the Virtual MET as an assessment tool for executive functions. *Neuropsychological Rehabilitation*, 19, 583–602.
- Reitan, R.M. (1992). *Trail Making Test: Manual for administration and scoring*. Tuscon, AZ: Reitan Neuropsychology Laboratory.
- Rempfer, M.V., Hamera, E.K., Brown, C.E., & Cromwell, R.L. (2003). The relations between cognition and the independent living skill of shopping in people with schizophrenia. *Psychiatry Research*, 117, 103–112.
- Robbins, T.W. (1990). The case for frontostriatal dysfunction in schizophrenia. *Schizophrenia Bulletin*, 16, 391–402.
- Schelling, D. (1993). *Corsi-Block-Tapping-Test*. Mödling, Austria: Schuhfried.
- Semkovska, M., Bédard, M.-A., Godbout, L., Limoge, F., & Stip, E. (2004). Assessment of executive dysfunction during activities of daily living in schizophrenia. *Schizophrenia Research*, 69, 289–300.
- Shallice, T., & Burgess, P.W. (1991). Deficits in strategy application following frontal lobe damage in man. *Brain*, 114, 727–741.
- Startup, M., Jackson, M.C., & Bendix, S. (2002). The concurrent validity of the Global Assessment of Functioning (GAF). *British Journal of Clinical Psychology*, 41, 417–422.
- Ufer, K. (2000). *BADS: Behavioural Assessment of Dysexecutive Syndrom (German Version)*. Göttingen, Germany: Hogrefe.
- Velligan, D.I., Bow-Thomas, C.C., Mahurin, R.K., Miller, A.L., & Halgunseth, L.C. (2000). Do specific neurocognitive deficits predict specific domains of community function in schizophrenia? *Journal of Nervous and Mental Disease*, 188, 518–524.
- von Aster, M., Neubauer, A., & Horn, R. (2006). *Wechsler-Intelligenztest für Erwachsene WIE [Wechsler Test of Intelligence for Adults WIE]*. Frankfurt, Germany: Harcourt Test Services.
- Wiedl, K.H., & Uhlhorn, S. (2006). *OAF-P: Osnabrücker Arbeitsfähigkeiten-Profil [OAF-P: Osnabrueck Work Ability Profile]*. Göttingen, Germany: Hogrefe.
- Wilson, B.A., Alderman, N., Burgess, P.W., Emslie, H., & Evans, J.J. (1996). *BADS: Behavioural Assessment of the Dysexecutive Syndrome*. Bury St Edmunds, UK: Thames Valley Test Company.
- Wilson, B.A., Evans, J.J., Emslie, H., Alderman, N., & Burgess, P. (1998). The development of an ecologically valid test for assessing patients with dysexecutive syndrome. *Neuropsychological Rehabilitation*, 8, 213–228.